Comparative Analysis Of Various Noise Types Using Empirical Mode Decomposition Based Hurst Exponent Techniques

Poovarasan Selvaraj, Dr.Chandra Eswaran

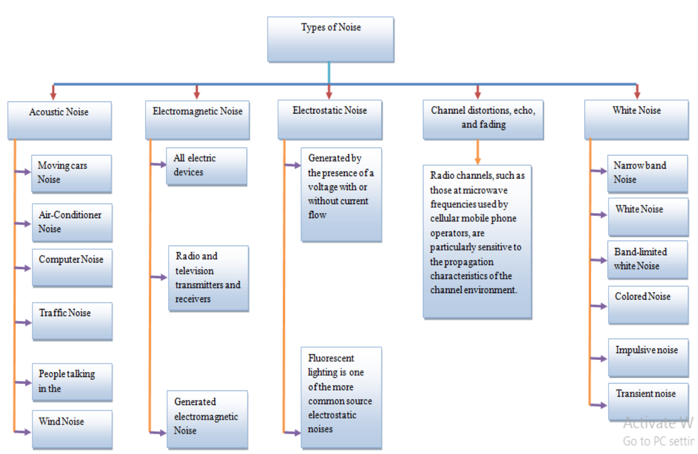
Abstract— Generally, Speech enhancement aims to develop a speech quality and intelligibility of a noise corrupted speech signal by using various Speech Enhancement techniques. Speech enhancement approach, Empirical Mode Decomposition and Hurst-based (EMDH) approach was proposed for signals corrupted by non-stationary acoustic noises. In this technique, Hurst exponent statistics was adopted for identifying and selecting the set of Intrinsic Mode Functions (IMF) that are most affected by the noise components. The results show that the EMDH improves speech quality were evaluated by the performance matrices of Cross Correlation, Mean Square Error, Peak Signal to Noise Ratio and the perceptual evaluation of speech quality (PESQ). An experimental study was also done on various types of noise added in clean speech like Gaussian White Noise, Random Noise and Colored Noise.

Index Terms— Empirical mode decomposition, Sifting Process, Hurst Exponent, Various types of Noise, EMDH.

1 INTRODUCTION

The enhancement of speech corrupted by noise is an important problem with number of applications ranging from control of environmental noise for communication systems and Voice over Internet Protocol, Hearing aids application, speech coding, mobile communications, and enhancing quality of old records, to preprocessing for speech recognition systems. Speech processing signals from the uninhibited environment may contain degradation components along with the essential speech components [1]. Degradation components include Reverberation, speech from other speakers, background noise. However, removing the irrelevant information must 2 Procedure for Paper Submission not degrade the relevant information. Figure 1 explains about the different types of noises [2].

Fig 1: Types of Noise



The different noise decibels are harmful in various ways to humans. The higher noise decibel can affect humans from minimum of hearing related problems to maximum of permanent hearing loss. If the humans are hearing Faint noise, Moderate noise and loud noise in more than over 85 dB for expected periods and it can cause them up to permanent hearing loss. Hearing the very loud noise for continuously or more than 30 minutes, then it is very dangerous to humans and it will cause serious hearing problems. Hearing the uncomfortable noise for continuously or more than 30 seconds, and then it is also very dangerous to humans. At last, the painful noise and dangerous noise which creates noise from 120 decibel to 130 decibels also results in more negative to the humans. To avoid these unwanted noises and also to protect the humans from the hearing related problems which were raised by these different noises, the hearing protection mechanism were introduced. And that hearing protection mechanism was adopted in many hearing devices to shelter the humans and to keep away from the hearing related problems raised by those different noise levels. Figure 2 illustrates the different noise's dB Levels.

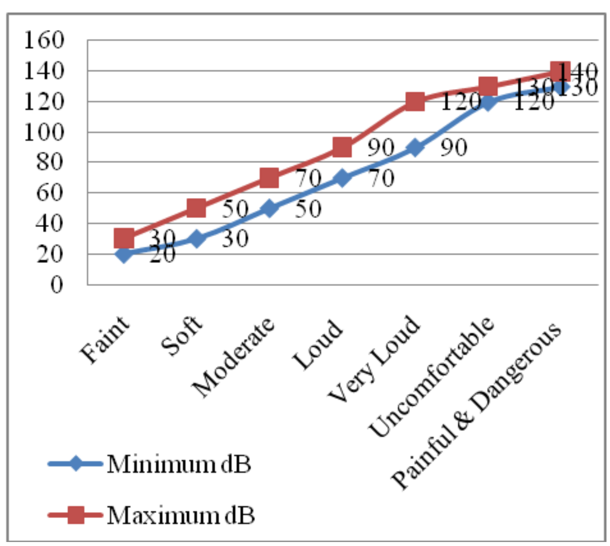


Figure: 2 Noise Level

- *Poovarasan Selvaraj, Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore 641 046,India Phone No:+91 9597942454, pppoovarasan@gmail.com*
- *Dr.Chandra Eswaran Professor Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore 641 046,India, crcspeech@gmail.com*

2. Literature Review

A colored noise based multi-condition training technique [3] was proposed for robust speaker identification in unknown noisy environments. In this technique, the colored noise samples generation was based on filtering a white Gaussian sequence. Gaussian Mixture Models (GMM) was applied for obtaining the speaker models by using the noisy speech signals with a single SNR. However, the identification accuracy was less.

EMD-based Filtering (EMDF) of low-frequency noise [4] was proposed for speech enhancement. In this technique, an adaptive method was developed for selecting the IMF index to separate the noise components from the speech according to the second-order IMF statistics. Then, the low-frequency noise components were separated by a partial reconstruction from the IMF. Based on this technique, a residual noise was suppressed from the speech signals that were enhanced by the conventional optimally modified log-spectral amplitude approach that utilizes a minimum statistics-based noise estimate. However, a minor improvement was required with the non-stationary Babble noise.

Speech enhancement strategy [5] was proposed based on time adaptive thresholding of IMF of the signal extracted by EMD. The denoised signal was reconstructed by the superposition of its adaptive threshold IMFs. The adaptive thresholds were estimated by using the Teager–Kaiser energy operator (TKEO) of signal IMFs. It was used to identify the type of frame by expanding differences between speech and non-speech frames in each IMF. However, the parameters used for implementing a compromise between noise removal and speech distortion were required to optimize for further improvement.

3 EMDH Methodology

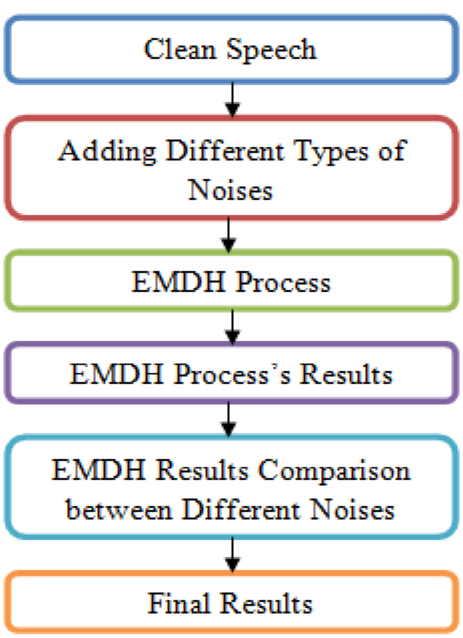


Figure 3: EMDH Methodology

Figure 3 shows the methodology of the work which was adopted and experimented in this research paper.

3.1 Empirical Mode Decomposition

EMD is techniques basically use to break down signals without leaving the time domain. It can be evaluated and compared to other analysis methods like Fourier Transforms and wavelet decomposition. The procedure is useful for analyzing natural signals, which are non-linear and non-stationary in nature [2]. The EMD method is a vital step to minimize any given data into a collection of intrinsic mode functions (IMF) to which the Hilbert spectral analysis can be initially started. The following requirements should be considered to satisfy the IMF [10]

1. In the whole data set, the number of extrema and the number of zero-crossings should either be equal or differ by one mostly.
2. At any point, the mean value of the envelope mentioned by the local maxima and the envelope defined by the local minima is zero.

The EMD Algorithm: The sifting process can be summarized in the following algorithm. Decompose a data set y(d) into IMFs yn(d) and a residuum r(d) such that the signal can be represented as:

$$y(d) = \sum_{n} y_n(d) + r(d)$$

3.2 Stopping the Sifting Process

In accordance with the second criteria of IMF, the mean of its envelope is equivalent to zero at each point of IMF. Therefore, a termination criterion is used for the sifting process. In each iteration, the ratio of the mean value of the envelope of iterated mode and the amplitude of this envelope is verified.

$$\tau(t) = \left|\frac{m(t)}{a(t)}\right|$$

$$where\ a(t) = \frac{e_{max}(t) - e_{min}(t)}{2}$$

Sifting process [6] is terminated if $\tau(t) < \beta_1$ is true for $(1-\mu)$ part of the number of signal's points and if $\tau(t) < \beta_2$ is true for the remaining points. The typical values of these parameters are given as:

$$\mu = 0.05;\ \beta_1 = 0.05;\ \beta_2 = 10 \cdot \beta_1$$

3.3 Hurst-based IMF Selection

The EMD algorithm states that, if a speech signal y(i) is decomposed as in (1), its reconstructed using only a subset of the first M IMFs

$$x(i) = \sum_{n=1}^{M} IMF_n(i).with\ M < N$$

The main issue [7] of the EMDH approach is the taking on the Hurst Exponent to make a decision which on IMFs must be preferred for the speech signal decomposition. Consider the speech signal be represented by a stochastic method y(i), among the normalized autocorrelation coefficient function ($\alpha(j)$) as

$$\alpha(j) = \frac{E[(y(i) - \mu_x)(y(i+j) - \mu_x)]}{E[(y(i) - \mu_x)^2]}$$

In equation, μ_x refers the mean of y(i) and j referes the time lag. For a fractional Gaussion Noise, $\alpha(j)$ is given as:

$$\alpha(j) = \frac{1}{2}(|i-1|^{2H} - 2|i|^{2H} + |i+1|^{2H})$$

The Hurst Exponent is expected from non-overlapping frames of samples and it used to enable the sorting criteria for selecting the IMF low frequency noise [8].

4. Experimental Results

In this section, performance effectiveness of the EMDH is evaluated and compared with different types of noise using by MATLAB 2014a. In this experiment, subsets of 12 speakers of clean speech and added Noise for Gaussian White Noise, Random Noise, and Colored Noise. The following are the performance metrics used to evaluate the effectiveness of the EMDH technique:

	Performance matrix	Range
1	Cross Correlation	0-1
2	Mean Square Error	0-1
3	Peak Signal to Noise Ratio	0-100 (in dB)
4	Perceptual Evaluation of Speech Quality	0.5-4.5

Table 1: Ranges of performance Matrices

4.1 Cross Correlation

It is a measure that gives the correlation or similarity score between related elements in the original speech signal. It computed as follow:

$$r_{xy} = \frac{c_v(x,y)}{\sigma_x \sigma_y}$$

4.2 Mean Square Error (MSE)

It represents the cumulative squared error between the reconstructed and original speech signal. The MSE is calculated as:

$$MSE = \frac{1}{l}\sum_{i=1}^{n} e_i^2 \ \ where\ e = \hat{x}(t) - x(t)$$

4.3 Peak Signal-to-Noise Ratio (PSNR)

It is defined as the fraction of maximum possible signal power to the corrupting noise power. Generally, it is computed by using MSE as:

$$\mathrm{PSNR(dB)} = 10\log_{10}\frac{255^2}{\mathrm{MSE}}$$

4.4 Perceptual Evaluation of Speech Quality (PESQ)

It can be applied to provide an end-to-end quality assessment for characterizing the listening quality as perceived by users [11].

$$\mathrm{PESQ} = x_0 - x_1.D - x_2.A$$

$$\text{Where } x_0 = 0.1, x_1 = 0.1, \text{and } x_2 = 0.0309$$

Table 2: Comparison result for Types of Noise

Different types of Noise	CC	MES	PSNR	PESQ
Gaussian White Noise	0.773815	0.000781	21.430105	2.485616
Random Noise	0.634490	0.001661	20.719679	2.577601
Colored Noise	0.707492	0.001394	18.850990	2.302364

Table 1 shows the comparison of different parameters between the three noises. And also those results were shown in chart formats. Figure 04 explains the Cross Correlation comparison between three noises. Figure 05 illustrates the MSE comparison between three noises.

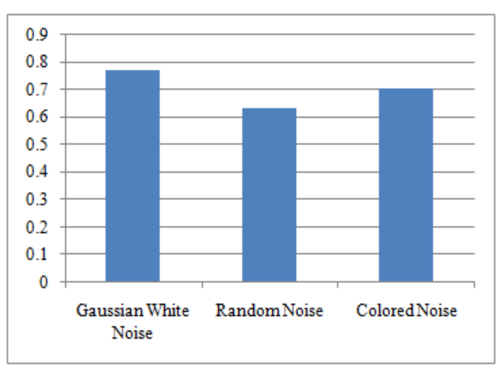


Figure: 3 Comparison of Cross Correlation

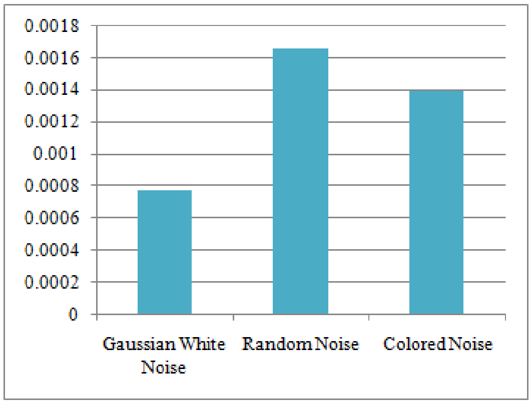


Figure: 4 Comparison of Mean Square Error

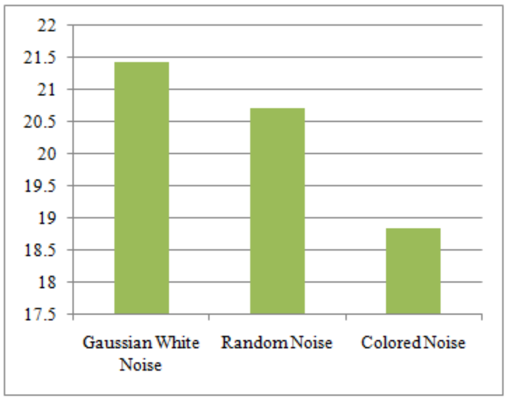


Figure: 5 Comparison of Peak Signal to Noise Ratio

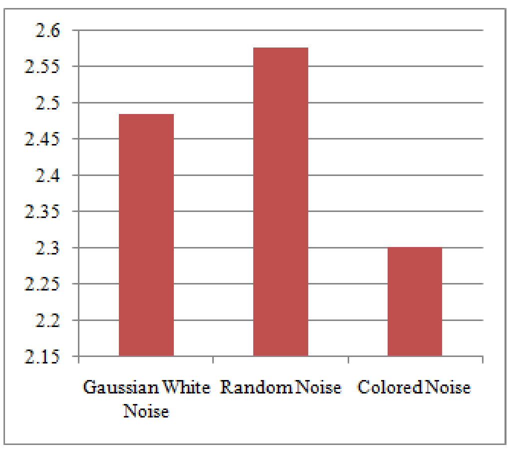


Figure: 6 Comparison of Perceptual Evaluation of Speech Quality

Figure 06 explains the PSNR comparison between three noises. Figure 07 illustrates the PESQ.

5. Conclusion

In this research article, the proposed methodology introduced a new Speech enhancement approach based on EMD and on a Hurst-based IMF Selection. And those proposed approach was tested using the three different types of noises. The outcome of the proposed methodology's results were evaluated and compared with the common parameters which that include Cross Correlation, MSE, PSNR and PESQ. Among those tested noises the proposed methodology technique is better. In future, the other noises will be tested by using the proposed methodology technique.

Reference

[1]. Alugonda Rajani et.al "A Review on Various Speech Enhancement Techniques" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 8, August 2016

[2]. Navneet Upadhyay et.al "Single Channel Speech Enhancement: using Wiener Filtering with Recursive Noise Estimation" Procedia Computer Science 84 (2016) 22 – 30

[3]. L. Zão and R. Coelho, "Colored noise based multicondition training technique for robust speaker identification," IEEE Signal Process. Lett., vol. 18, no. 11, pp. 675–678, Nov. 2011

[4]. Chatlani, N., & Soraghan, J. J. (2012). EMD-based filtering (EMDF) of low-frequency noise for speech enhancement. IEEE Transactions on Audio, Speech, and Language Processing, 20(4), 1158-1166.

[5]. Khaldi, K., Boudraa, A. O., & Komaty, A. (2014). Speech enhancement using empirical mode decomposition and the Teager–Kaiser energy operator. The Journal of the Acoustical Society of America, 135(1), 451-459.

[6]. Dwijayanti, S., Yamamori, K., & Miyoshi, M. (2018). Enhancement of speech dynamics for voice activity detection using DNN. EURASIP Journal on Audio, Speech, and Music Processing, 2018(1), 10.

[7]. Ghahabi, O., Zhou, W., & Fischer, V. (2018). A Robust Voice Activity Detection for Real-Time Automatic Speech Recognition. In Proceedings of ESSV 2018, Ulm 2018.

[8]. Chatlani, N., & Soraghan, J. J. (2012). EMD-based filtering (EMDF) of low-frequency noise for speech enhancement. IEEE Transactions on Audio, Speech, and Language Processing, 20(4), 1158-1166

[9]. Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An algorithm for intelligibility prediction of time–frequency weighted noisy speech. IEEE Transactions on Audio, Speech, and Language Processing, 19(7), 2125-2136.

[10]. Mai, V. K., Pastor, D., Aïssa-El-Bey, A., & Le-Bidan, R. (2015). Robust estimation of non-stationary noise power spectrum for speech enhancement. IEEE Transactions on Audio, Speech, and Language Processing, 23(4), 670-682.

[11]. Md. Ekramul Hamid et.al " Speech Enhancement Using EMD Based Adaptive Soft-Thresholding (EMD-ADT)"International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 5, No. 2, June, 2012.